

LE PROJET LeKo :
**UNE RECHERCHE PHRASÉOLOGIQUE
INTÉGRALE POUR L'APPRENTISSAGE DES
PHRASÈMES EN CONTEXTE PLURILINGUE**

Lorenzo Zanasi
EURAC RESEARCH (ITALIA)

lorenzo.zanasi@eurac.edu

<https://orcid.org/0000-0002-4439-6567>

Christine Konecny
UNIVERSITÄT INNSBRUCK (ÖSTERREICH)

Christine.Konecny@uibk.ac.at

<https://orcid.org/0000-0002-7469-1225>

Erica Autelli
UNIVERSITÄT INNSBRUCK (ÖSTERREICH)

Erica.Autelli@uibk.ac.at

<https://orcid.org/0000-0003-4473-4252>

Andrea Abel
EURAC RESEARCH (ITALIA)

andrea.abel@eurac.edu

<https://orcid.org/0000-0003-4694-0295>

Résumé : Dans cet article, on décrit la méthodologie et les résultats d'un projet de recherche triennal (2013-2016) consacré à l'analyse et à l'enseignement des collocations et des phraséologismes de l'italien (LeKo). Résultant d'une collaboration entre l'Université d'Innsbruck et De Eurac Research, le projet LeKo se situe au croisement de trois domaines : les études phraséologiques, en particulier allemandes et italiennes, la linguistique de corpus et la didactique des langues étrangères et secondes. L'objectif innovant du projet est le recueil de

données visant la réalisation d'instruments destinés spécifiquement à la phraséodidactique de l'italien pour des apprenants germanophones.

Mots-clés : phraséologie, phraséodidactique, corpus d'apprenants, collocations, italien langue seconde.

Abstract : This article outlines the methodology and results of a three-year research project (2013-2016) focusing on the analysis and the didactics of Italian collocations and phrasemes (LeKo). LeKo, a joint project between the University of Innsbruck and Eurac Research, is particularly noteworthy as it combines phraseological studies (with an emphasis on German and Italian) with corpus linguistics and the teaching of foreign and second languages. LeKo's main innovative goal is to collect data to develop teaching material on Italian phraseodidactics for German-speaking learners.

Keywords : phraseology, phraseodidactics, learner corpora, collocations, Italian as a second language.

Introduction

Cet article s'inscrit dans le contexte de recherches en phraséologie et en phraséodidactique. Il s'agira ici de décrire une recherche dans laquelle la contribution de la linguistique de corpus à la phraséologie occupe une place centrale. Au cours de la contribution, nous verrons comment l'extraction des informations à partir de bases de données et l'analyse des aspects morphosyntaxiques produits par les apprenants, jouent un rôle important dans l'identification des éléments phraséologiques d'un texte. Les phénomènes phraséologiques, surtout les moins "voyants", c'est-à-dire les non idiomatiques, sont en effet hétérogènes et rarement visibles dans le discours, écrit ou oral. La méthodologie qui sous-tend la construction d'un corpus permet d'organiser les catégories phraséologiques, de les annoter puis de les retrouver avant d'y effectuer des observations quantitatives et des évaluations qualitatives. Il s'agit donc d'une méthodologie qui peut être d'une grande aide et d'un grand soutien dans l'enseignement des phraséologismes italiens à un public d'apprenants de l'italien, langue seconde ou étrangère.

Dans les paragraphes suivants, nous présenterons le projet LeKo, en commençant par une brève description du corpus d'apprenants (*learner corpora*) et de la méthodologie appliquée dans la recherche (section 2). Le schéma de classification des phénomènes phraséologiques étudiés et le schéma d'annotation (sections 3 et 4) seront également décrits. Puis, nous présenterons les principaux résultats extraits de l'analyse du corpus, en nous concentrant sur certains aspects de l'erreur et du transfert linguistique qui peuvent être pertinents à des fins didactiques (section 5) et en concluant par les évaluations finales (section 6). Dans cette contribution, nous nous focaliserons donc essentiellement sur la méthodologie adoptée pour la création du corpus LeKo et sur des phénomènes extraits du corpus, sur lesquels les objectifs pédagogiques pourraient se concentrer. L'aspect phraséodidactique au sens strict, deuxième axe du projet, sera remis à d'autres publications (Schmiderer *et al.*, à paraître ; Schmiderer *et al.*, à paraître).

La terminologie employée dans ces pages mérite d'être précisée. Le terme "phraséologisme" est utilisé ici dans son sens le plus large et le plus général, et non dans son sens le plus étroit⁴⁰. "Phrasème", d'après la terminologie de Burger (2003), est un mot employé comme synonyme de "phraséologisme". Dans sa conception large, le couple "phraséologisme"/"phrasème" se rapproche de l'expression anglaise *formulaic sequences*⁴¹ (à partir de maintenant : FS), qui se réfère à une large gamme d'expressions (voir Wray et Perkins, 2000).

2. Le projet LeKo

2.1 Le contexte de la recherche

LeKo est un projet scientifique d'une durée de trois ans (2013-2016), financé par la Province Autonome de Bolzano⁴², plus connue en français comme Haut-Adige ou Tyrol du Sud. Il est le fruit d'une coopération entre l'Institut de Romanistique de l'Université d'Innsbruck (responsable du projet) et l'Institut de Linguistique Appliquée de Eurac Research.

La recherche a été menée dans un contexte social et géographique caractérisé par les contacts de langues : l'allemand en Autriche et dans le Haut-Adige/Tyrol du Sud et l'italien dans le Haut-Adige/Tyrol du Sud. En plus de constituer une L1 pour les communautés locales respectives, l'allemand et l'italien sont également appris et enseignés en tant que L2 dans le cadre des échanges culturels, économiques et sociaux entre l'Autriche et l'Italie et en particulier en raison du régime de bilinguisme administratif dans le Haut-Adige/Tyrol du Sud. Le secteur phraséologique est donc d'une grande importance dans un tel contexte, où la communication interpersonnelle est fréquemment exposée à des pratiques de contact et de changement de code dans le répertoire linguistique individuel des locuteurs.

L'acronyme LeKo⁴³ qui identifie le projet est issu de l'allemand et articule principalement des termes comme "lexema" (*Lexem*) et "combinaisons" (*Kombinationen*), mais on peut également y attribuer des valeurs similaires comme "vocabulaire", "lexicologie", "contexte", "corpus", "linguistique de corpus", "collocations" et "construction". Il s'agit d'un domaine et de phénomènes hétérogènes qui, dans LeKo, contribuent à répondre aux questions clés de la recherche : quelles combinaisons de mots les étudiants germanophones du Tyrol du Sud utilisent-ils pour écrire en italien ? Quelles sont les erreurs les plus fréquentes qu'ils commettent ?

40 Voir l'entrée *Glossario* dans Konecny et Autelli (2012). Pour la conception large du terme "phraséologisme", voir par. ex. Fleischer (1982) et Dobrovolskij (2002). Pour la conception étroite, voir par. ex. Skytte (1988) et Palm (1997).

41 Définie comme « [...] a sequence, continuous or discontinuous, of words or other elements, which is [...] prefabricated [...] » (Wray, 2010: 9).

42 Financement accordé par la Division Droit d'étude, recherche universitaire et scientifique, loi provinciale du 13 décembre 2006, n° 14, Recherche et innovation.

43 Le titre complet du projet est le suivant : *Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext* (<http://www.leko-project.org/>).

Pour répondre à ces questions, le groupe de recherche a entamé une analyse linguistique basée sur la constitution d'un corpus dédié, qui conduit à la création de supports phraséodidactiques.⁴⁴

2.2 La base des données et la méthodologie de la recherche

Le corpus Leko est constitué d'un recueil de textes écrits en italien, rédigés par des apprenants germanophones. Les textes n'ont pas été créés spécifiquement pour le projet, mais proviennent d'un patrimoine textuel plus large qui appartient à Eurac Research et est le résultat de deux recherches précédentes : le projet Kolipsi⁴⁵ et le projet Merlin⁴⁶. Bien que poursuivant des objectifs différents, ils partagent des caractéristiques communes ; dans les deux projets, les productions textuelles ont été collectées, évaluées et assignées à un niveau du *Cadre européen commun de référence pour les langues* (désormais : CECRL). En outre, tous les textes sont accompagnés d'informations extralinguistiques : les méta-données des apprenants (niveau du CECRL, L1, L2, sexe, type d'école, zone géographique d'origine).

Le corpus LeKo est donc composé de textes de Kolipsi (290 productions écrites en italien L2, 144 étudiants du Tyrol du Sud de langue maternelle allemande, 64.651 occurrences) et de Merlin (50 productions écrites en italien L2 de 50 étudiants germanophones, 8.815 occurrences). On peut constater que même si le corpus n'est pas très grand, il est, à notre connaissance, le premier de son genre ; deuxièmement, même si le corpus en soi ne permet pas de tirer des conclusions générales, ses résultats peuvent aider les enseignants à trouver des exemples et à avoir une vision large des problèmes liés aux phénomènes phraséologiques de l'italien L2.

D'un point de vue méthodologique, il convient tout d'abord de souligner que, dans le projet LeKo, la sélection des phénomènes phraséologiques n'a pas eu lieu, comme cela aurait été possible, en créant des répertoires, ou des listes, sur la base de critères morphosyntaxiques, ni même en utilisant des procédures automatiques basées sur la fréquence des co-occurrences lexicales et sur celles qui, en statistique, sont appelées "mesures d'association" (par exemple, Sketch Engine). Au contraire, les phraséologismes objets d'analyse ont été extraits manuellement des productions textuelles des apprenants, productions caractérisées par des finalités communicatives précises et par un processus de création de sens dans un contexte. Cette origine des données signifie qu'une part considérable des occurrences recensées contient des erreurs, des traces interlinguistiques, des interférences ; tous ces éléments sont précieux d'un point de vue didactique, surtout lorsqu'ils sont croisés avec les méta-données des apprenants.

Une fois sélectionnés, les textes ont été analysés selon une combinaison de méthodes déductives (l'utilisation de la littérature scientifique pour classer les phénomènes) et de méthodes inductives (l'analyse empirique des textes écrits qui ont conduit à l'annotation).

44 Voir Schmidereret al. (2021).

45 Voir Abel, Vettori et Wisniewski (2017).

46 Voir Abel *et al.* (2013).

En ce qui concerne l'approche phraséologique, il convient de noter qu'elle a été interprétée dans une perspective lexicologique large et que de nombreuses entités phraséologiques ont donc été incluses dans l'analyse : des combinaisons semi-libres, collocations lexicales, des structures composées et des expressions idiomatiques.

En ce qui concerne l'approche computationnelle, les textes des corpus de Kolipsi et Merlin ont d'abord été étiquetés automatiquement à l'aide de Treetagger (Schmid, 1994), un instrument qui assigne des lemmes et des parties du discours (POS) dans un texte. Ensuite ces étiquettes ont été corrigées manuellement. L'annotation de texte a été réalisée à l'aide du logiciel MMAX2, « highly customizable tool for creating, browsing, visualizing and querying linguistic annotations on multiple levels » (Müller et Strube, 2006 : 198). Les FS ont été annotés en exploitant le potentiel du logiciel pour enrichir le corpus d'informations sur les erreurs, les formes cibles, les structures syntaxiques. La plateforme ANNIS (Zeldes *et al.*, 2009) a été utilisée pour interroger le corpus et afficher les données.

Enfin, un contrôle de qualité a été implémenté pour assurer la cohérence interne :

- (i) un manuel d'annotation détaillé pour les annotateurs,
- (ii) un échantillon d'annotations partagées et testées à prendre comme référence commune,
- (iii) l'annotation conjointe⁴⁷ d'environ cinq pour cent des textes, de sorte que la fiabilité des annotations puisse être calculée. Sur cette base, les premières analyses quantitatives ont été effectuées.

3. Le schéma de classification

Les phraséologismes analysés dans le cadre du projet Leko couvrent un large éventail de phénomènes, organisés en trois grandes classes : les phrasèmes référentiels, les phrasèmes communicatifs et les phrasèmes structuraux. Cette répartition et la terminologie utilisée sont fondées sur la classification mixte de Burger (2003 : 36 et suiv.). De plus, au sein de ces trois macro-groupes, on distingue des sous-classes, incluant ainsi un grand nombre de phénomènes, comme décrit ci-dessous.

Les phrasèmes référentiels (voir Fig. 1 et Fig. 2) ont été divisés en deux types (non-idiomatiques et idiomatiques), selon le degré d'idiomaticité de la séquence des lexèmes. Les deux typologies comprennent à leur tour quatre combinaisons différentes : (a) les collocations et les expressions idiomatiques, (b) les composés nominaux, (c) les verbes syntagmatiques et (d) les expressions adverbiales. Le terme "collocation" est utilisé ici au sens large, car il comprend également des expressions proches des combinaisons libres, comme *andare al supermercato* ['aller faire les courses'], les structures avec support verbal comme *fare un confronto* ['faire une comparaison'],

⁴⁷ En fait, « inter-annotator agreement is a quantitative index for the degree of agreement between the annotations produced by two or more annotators on the same base data, using the same annotation manually » (Müller et Strube, 2003: 207).

les comparaisons phraséologiques⁴⁸ comme *fumare come un turco* [‘fumer comme un pompier’] et les expressions semi-idiomatiques comme *numero verde* [‘numéro vert’]. Un exemple d’expression idiomatique extraite du corpus LeKo est *dare una mano* [‘donner un coup de main’]. Les composés nominaux qui se comportent comme des collocations (par exemple *lenti a contatto* [‘lentilles de contact’]) et les mots composés idiomatiques (par exemple *luna di miele* [‘lune de miel’]) sont souvent appelés “espressioni polirematiche” (Voghera, 2004), “composti polirematici” (Cicalese, 1995) ou “lexèmes complexes” (Voghera, 1994 ; De Mauro et Voghera, 1996). Les verbes syntagmatiques (voir Simone, 1997) comprennent des expressions de composition comme *buttare via* [‘jeter’] ou *pensarci sopra* [‘y penser’] ou, si elles sont idiomatiques, comme *passar su* [‘passer sur’]. Parmi les expressions adverbiales non-idiomatiques, on trouve des formes comme *in ogni caso* [‘en tout cas’] et aussi des formules binomiales comme *di anno in anno* [‘d’année en année’] ; parmi les expressions idiomatiques on trouve des formes comme *a occhio e croce* [‘à peu près’] et *in fin dei conti* [‘en fin de compte’].

Pour Burger, les phrasèmes communicatifs (voir Fig. 3) sont liés à des actes communicatifs et à des formules routinières qui ont pour tâche d’accomplir des fonctions pragmatiques. D’une part, nous incluons dans ce regroupement des expressions liées à la situation communicative comme des séquences récurrentes d’interactions interpersonnelles telles que les formules de salutation et de louange (*ciao, cordiali saluti* [‘salut’, ‘cordialement’]), les commentaires méta-communicatifs (*come dico sempre* [‘comme je dis toujours’]), qui peuvent exprimer l’importance de l’expression (*a moi avviso* [‘à mon avis’]), le flou (*per quanto ne so io* [‘d’après mon expérience’]), la surprise (*accipicchia!* [‘Oh mon Dieu!’]) et les formes binomiales qui remplissent des fonctions pragmatiques (*cincin* [‘santé’]). D’autre part, nous incluons des expressions qui ne sont pas liées à la situation communicative, mais qui facilitent la compréhension (*a dir la verità, intendo dire* [‘à vrai dire’, ‘je veux dire’]), qui guident l’attention (*come vedi* [‘comme tu vois’]), qui apportent des références internes de texte (*come detto* [‘comme indiqué’])⁴⁹.

Il y a aussi des phrasèmes structuraux (voir Fig. 4), qui n’ont pas seulement une fonction linguistique, mais aussi celle de créer des relations grammaticales (Burger, 2003 : 36). Parmi celles-ci, on peut distinguer les phrasèmes structuraux principalement composés de prépositions (*assieme a, in seguito a* [‘ensemble’, ‘à la suite de’]) ou principalement de conjonctions (*per quanto riguarda, sia...sia* [‘en ce qui concerne’, ‘tant... que’]).

48 Voir García-Page Sánchez (1996) et aussi Hausmann (2004). Le terme utilisé par Hausmann (2004 : 314) pour ces combinaisons est “Vergleichsphrase” (‘phrasèmes comparatifs’).

49 Pour la distinction entre les expressions liées ou non au contexte communicatif, voir Stein (2007).

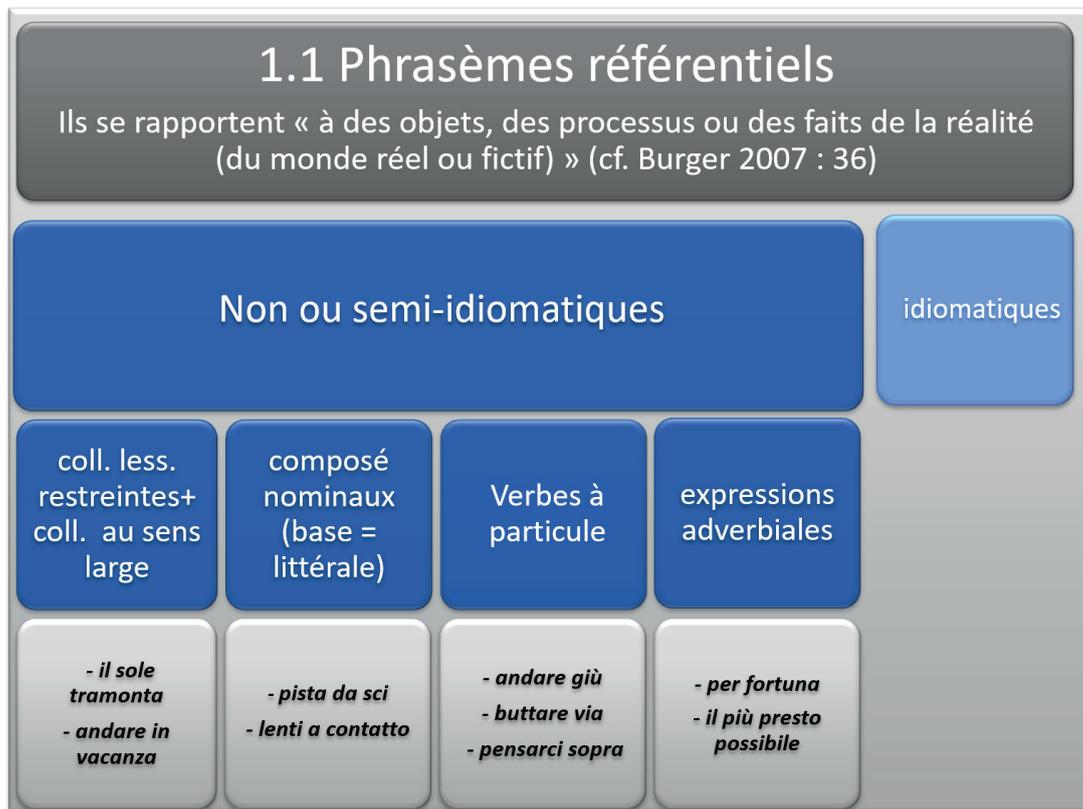


Fig. 1 : phrasèmes référentiels non ou semi-idiomatiques extraits du *corpus* LeKo (De gauche à droite : le soleil se couche – aller en vacances – piste de ski – lentilles de contact – descendre – jeter – y penser/y réfléchir – heureusement – au plus vite/dès que possible)

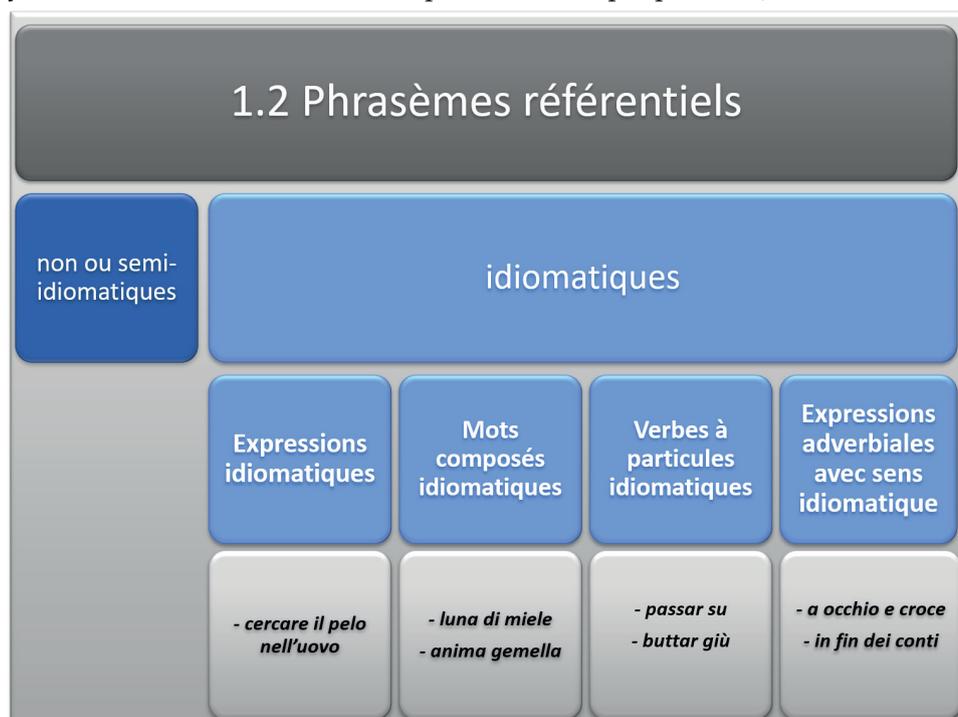


Fig. 2 : phrasèmes référentiels idiomatiques extraits du *corpus* LeKo (De gauche à droite : chercher midi à quatorze heures/chercher la petite bête – lune de miel – âme sœur – passer sur – jeter sur le papier – à peu près – en fin de compte)

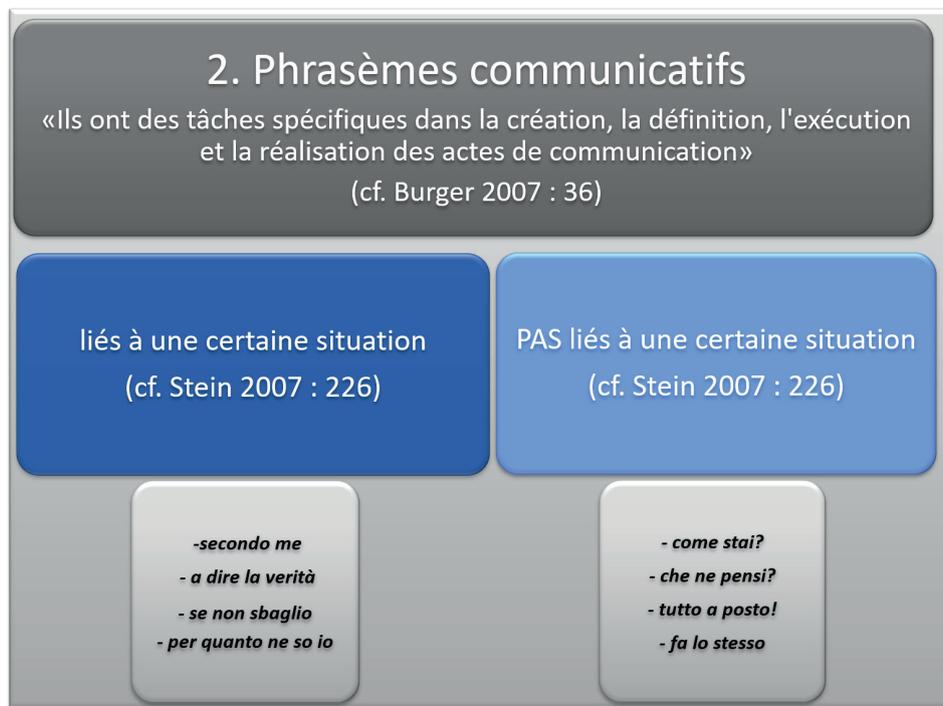


Fig. 3 : phrasèmes communicatifs extraits du *corpus* LeKo (De gauche à droite : selon moi – à vrai dire – si je ne me trompe pas – pour autant que je sache – comment ça va ? – qu’est-ce que tu en penses ? – tout va bien ! – ce n’est pas grave)

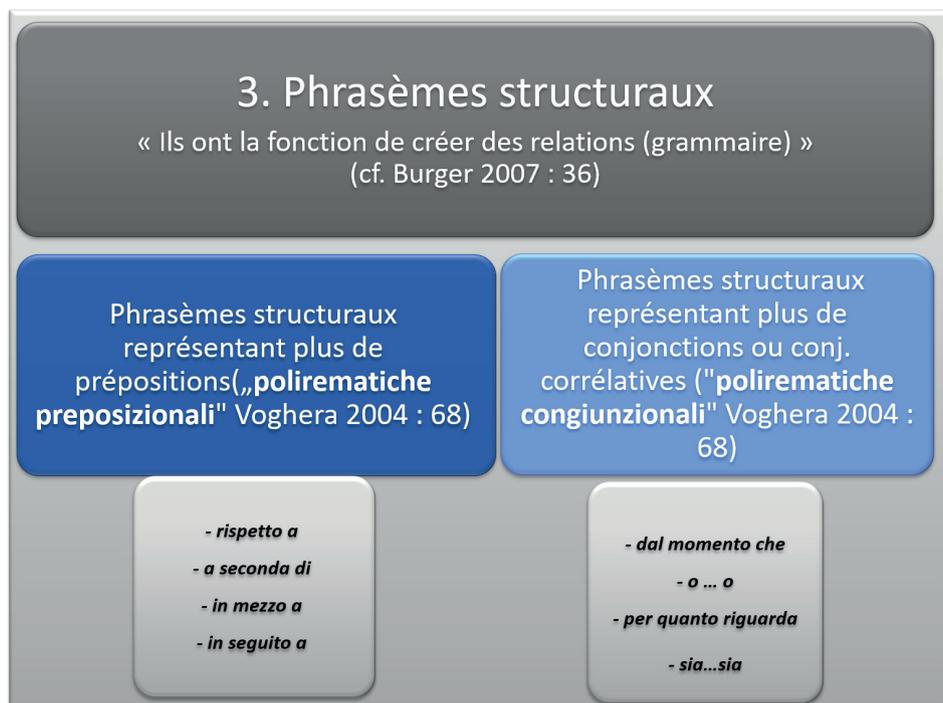


Fig. 4 : phrasèmes structuraux extraits du *corpus* LeKo (De gauche à droite : par rapport à – en fonction de – au milieu de – à la suite de – étant donné que – soit/soit – en ce qui concerne – que ce soit/ou bien)

Une dernière catégorie, particulièrement importante pour l'enseignement, est celle qui comprend des combinaisons qui sont clairement des traductions de l'allemand et qui n'existent pas en italien. L'annotation de ces séquences est utile pour deux raisons : les séquences contiennent parfois des informations culturelles qui expliquent le résultat de l'interférence et s'inscrivent dans le *continuum* du jugement d'acceptabilité d'un locuteur natif. C'est la raison pour laquelle elles sont étiquetées comme étant soit inexistantes, mais acceptables, soit non existantes et non acceptables.

4. Le schéma d'annotation

Pour la mise en œuvre et la consultation du corpus, un schéma d'annotation manuelle a été développé, caractérisé par la combinaison de différents niveaux d'analyse. À chaque niveau, on considère les phrasèmes sélectionnés à partir d'un point de vue différent, en suivant un schéma de marquage textuel (voir Fig. 5), composé de 64 marqueurs (tag) et divisé en quatre phases :

- l'identification des phrasèmes selon les cinq catégories illustrées au paragraphe précédent ;
- dans le cas des collocations, l'annotation de leur structure (nom-verbe, verbe-préposition-nom, nom-adjectif, etc.) et de la fonction syntaxique de certaines catégories ;
- l'annotation des erreurs. Les marqueurs d'erreurs couvrent de nombreux cas : l'erreur peut être sémantique ou formelle. Dans le premier cas, les occurrences erronées d'un point de vue dénotatif ou connotatif sont incluses. Dans le second cas, sont incluses les occurrences erronées dues à l'omission d'un ou de plusieurs éléments de la séquence, à l'ajout d'un ou de plusieurs éléments non prévus par la séquence, à la mauvaise position d'un ou de plusieurs éléments de la séquence, au mauvais choix d'un ou de plusieurs éléments de la séquence. Le choix peut être imprécis sur le plan grammatical (par exemple, accords manquants : **alto stagione* au lieu de *alta stagione* ['haute saison']), lexical (**far chiaro a* au lieu de *chiarire* ['préciser/clarifier']) ou orthographique (**mile grazie* au lieu de *mille grazie* ['merci beaucoup']).
- la transcription de la forme cible (*target hypothesis* [TH]). Si la FS annotée contient une erreur, il y a un champ dans lequel la forme correcte peut être transcrite (par exemple pour la FS *foglio di cassa*, on écrira *scontrino* ['ticket de caisse']).
- l'explication possible de l'erreur selon l'interférence inter-linguistique⁵⁰ (**magazzino di viaggio* au lieu de *rivista di viaggio* ou **moi marito aveva compleanno* au lieu de *moi marito compiva gli anni* ['mon mari a fêté son anniversaire], transfert de l'allemand *mein Mann hatte Geburtstag*) et intralinguistique (**farsi la spesa* au lieu de *fare la spesa* ['faire les courses'] ; possible analogie sur le modèle *farsi una pizza, farsi una birra*, dans lequel la forme pronominale est utilisée).

⁵⁰ Voir Díaz-Negrillo et Fernández-Domínguez (2006), et aussi Lüdeling *et al.* (2005).

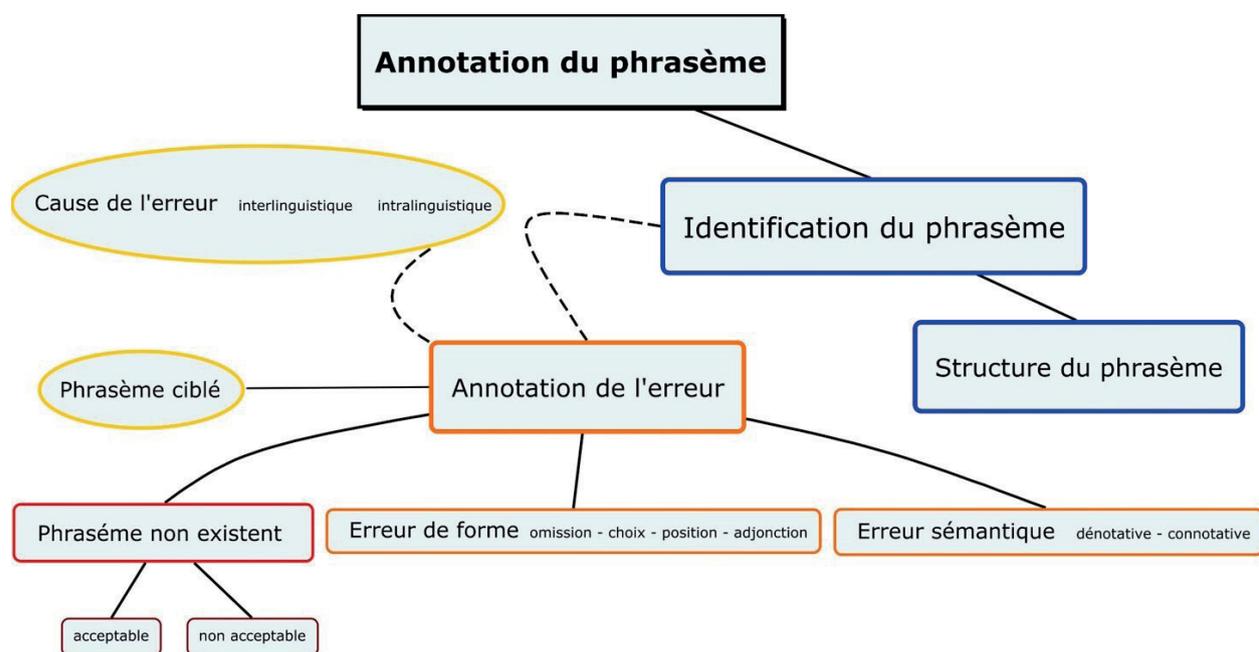


Fig. 5 : schéma d'annotation des FS.

5. Résultats de l'analyse du corpus

Nous allons maintenant illustrer brièvement certaines des conclusions tirées au cours de l'analyse du corpus.

En général, les éléments de combinaison complexes tels que les FS sont plus fréquents et plus corrects chez les apprenants de niveaux avancés du CECRL. L'analyse des données montre comment le pourcentage de FS utilisé par les apprenants augmente pour les niveaux B2/C1 par rapport aux niveaux A2/B1.

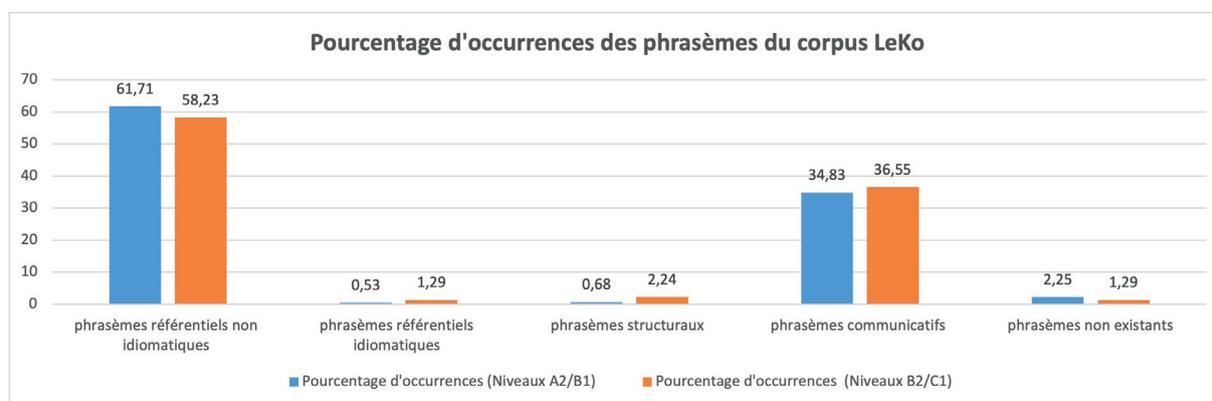


Fig. 6 : pourcentage de différences dans la présence des phrasèmes entre les niveaux inférieurs et supérieurs du CECRL.

L'exactitude et la fréquence sont directement proportionnelles à l'activation de la *expectancy grammar*, pour laquelle les structures prévisibles sont utilisées plus correctement que les structures imprévisibles (voir Balboni, 2016 : 170 et suiv.). Cela implique que les séquences complètement idiomatiques sont utilisées par des apprenants qui ont confiance en leurs compétences, qui sont sûrs de ne pas faire d'erreurs et de produire la bonne séquence au bon moment ; les expressions idiomatiques sont donc le plus souvent absentes chez les apprenants moins compétents.

Les catégories les plus représentées parmi les trois classes de phrasèmes sont les phrasèmes référentiels non-idiomatiques (particulièrement les collocations sans restriction, comme *andare in ferie* ou *andare al supermercato*) et les phrasèmes communicatifs (surtout les formes attribuables aux salutations et aux formules de politesse). La raison de la forte présence de ces catégories est liée au type de texte (courriel) et à la tâche de communication exigée des étudiants. Fait significatif, ces catégories figurent également dans les textes des niveaux inférieurs du CERCL.

En tant que tel, le corpus des apprenants contient de nombreuses occurrences caractérisées par des erreurs. Cependant, une production phraséologique correcte prévaut : le taux de FS sans erreur est en fait supérieur à 60% pour les niveaux A2/B1 et supérieur à 70 % pour les niveaux B2/C1. L'erreur se produit d'abord comme une erreur formelle de choix dans la mise en œuvre des SF, alors qu'il y a beaucoup moins d'erreurs sémantiques. Les erreurs de transfert ne représentent qu'un faible pourcentage des événements et, dans ce pourcentage, il y a une prépondérance absolue de l'interférence interlinguistique sur l'interférence intralinguistique. Cependant, les élèves du Tyrol du Sud affichent des comportements linguistiques influencés par la langue maternelle allemande de manière réduite, même s'ils sont systématiques et donc significatifs d'un point de vue éducatif. Enfin, le croisement avec certaines métadonnées montre comment, au lycée, le taux d'erreur diminue par rapport aux établissements techniques et comment les filles utilisent plus de phrasèmes communicatifs que les garçons.

Ainsi, les résultats extraits du corpus restituent une photographie dans laquelle l'interférence tend à se fossiliser dans certaines structures ; dans ce scénario, les compétences et les performances sont directement proportionnelles. Plus le niveau de compétence atteint est élevé, plus l'attention portée au niveau linguistique est grande, et plus les FS sont utilisées (et plus l'utilisation correcte est importante).

En approfondissant l'analyse et en extrayant l'information qui peut avoir un effet d'entraînement au niveau didactique, on peut observer différents cas⁵¹:

- erreurs d'omission (**ci sentiamo settimana prossima !* → TH : *ci sentiamo la settimana prossima !* ['à la semaine prochaine!'] ;
- erreurs par adjonction (**come ti va?* → TH : *come va ?* ['comment ça va ?']) ;
- erreurs de position (**prego mi scrivi* → TH : *per favore scrivimi* ['s'il te plaît, écris-moi'] ; **cosa raccontarmi ?* → TH : *cosa mi racconti ?* ['qu'est-ce que tu me racontes ?']). Le contrôle de la position des éléments phraséologiques est également critique dans les collocations avec

51 Les exemples montrent la forme italienne extraite du corpus, la forme cible (TH) et l'équivalent français. En cas d'interférence, la séquence est également transcrite en allemand.

une séquence nom-adjectif (*bel/bello*: **avere un tempo bello* → TH : *avere un bel tempo* [‘avoir du beau temps’] ; **andare in montagna alta* → TH : *andare in alta montagna* [‘aller en haute montagne’]).

- Au niveau de l’erreur sémantique, on peut indiquer les cas suivants:
- l’utilisation des FS existants, mais dans le mauvais contexte ou avec un sens erroné: **io ho pensato un po’ e mi sono messo d’accordo che* [‘se mettre d’accord avec qqn.’ dans le sens de ‘prendre une décision’] → TH : *sono giunto alla conclusione che* ; **far sicuro che* (transfert de l’allemand *sicher stellen, dass*) → TH : *assicurarsi che* [‘s’assurer que’] ;
- l’utilisation des FS qui n’existent même pas en italien: **una vacanza che ci teniamo in ricordo* (transfert de l’allemand *ein Urlaub, den wir in Erinnerung behalten*) → TH : *una vacanza che ci rimane in mente* [litt. ‘des vacances qui restent gravés dans la mémoire’].

Un phénomène d’interférence récurrent émerge dans la sur-extension du modèle syntagmatique verbal, dont l’allemand fait largement usage, alors qu’en italien l’adéquation de certaines expressions (¹*la commessa l’ha chiamata indietro* [‘la vendeuse l’a rappelée’] ; ²*ha messo tutte le cose che aveva comprato indietro* [‘elle a rendu tout ce qu’elle avait acheté’] ; ³*quando mi sono tirato via* (cf.) l’allemand *als ich mich zurück gezogen habe*) [‘expression mal utilisée au sens de ‘je suis sorti’] n’est pas acceptable ou apparaît uniquement dans des contextes spécifiques.

Le contrôle des opérateurs prépositionnels est également incertain : dans les phrasèmes structuraux, il y a des expressions de localisation incorrectes (**in mezzo della natura*, **in mezzo dei fiori* → TH : *in mezzo alla natura/ai fiori* [‘au milieu de la nature’, ‘au milieu des fleurs’]) et des formes incorrectes peuvent être trouvées dans les rections verbales (**aspetto alla tua risposta* [cf. l’allemand *ich warte auf deine Antwort*] → TH : *aspetto la tua risposta* [‘j’attends ta réponse’]) et dans le choix du bon opérateur (préposition) au sein des collocations (**andare nelle ferie* → TH : *andare in ferie* [‘partir en vacances’] ; **non ero in fretta* → TH : *non ero di fretta* [‘être pressé’]).

Ces dernières erreurs sont fréquentes dans les productions évaluées aux niveaux inférieurs du CECRL. Enfin, il y a la présence de combinaisons non-existantes qui sont le fruit d’une interférence de l’allemand, mais qui sont acceptables d’un point de vue sémantique (**prendere un bagno di sole* [transfert de l’allemand *ein Sonnenbad nehmen*] → TH : *abbronzarsi* [‘se faire bronzer’]).

Ceux-ci incluent souvent des occurrences avec le verbe *ricevere* [‘recevoir’], en raison de l’interférence du verbe allemand *bekommen* :

- **riceviamo anche la colazione* (cf. l’allemand *wir bekommen auch ein Frühstück*) → TH : *ci darebbero anche la colazione* [‘ils nous donneraient même le petit déjeuner’] ;
- **ricevere una crisi di nervi* (cf. l’allemand *einen Nervenzusammenbruch bekommen*) → TH : *avere una crisi di nervi* [‘faire une dépression nerveuse ; avoir une crise de nerfs ; piquer une crise de nerfs’] ;
- **ricevere la ragione* (cf. l’allemand *Recht bekommen*) → TH : *avere ragione* [‘avoir raison’] ;
- **abbiamo ricevuto un bell colore* (cf. l’allemand *wir haben eine gute/schöne Farbe bekommen*) → TH : *abbiamo preso un bel colorito* [‘nous avons pris des couleurs’] ;

- et ceux qui, en italien, ne correspondent pas à une FS, mais à un seul mot :
- **penso che noi abbia divertimento* (cf. l'allemand *ich denke, dass wir Spaß haben werden*) → TH : *penso che ci divertiremo* ['je pense qu'on va bien s'amuser'] ; **fai le tue pensiere* (*mach dir Gedanken darüber*) → TH : *pensaci* ['réfléchis'] ;
- **si sono fatta già pensieri* (cf. l'allemand *sie hat sich schon Gedanken darüber gemacht*) → TH : *si è messa a pensare/ha già cominciato a pensare a...*['elle a commencé à penser à...'] ;
- **così lo ho dentro nella mia testa* (cfr. l'allemand *so habe ich es im Kopf*) → TH : *così me lo immagino* ['comme je l'imagine'].

En dernier lieu, les séquences qui sont sémantiquement compréhensibles, mais construites selon le modèle combinatoire allemand sont celles qui présentent :

- (1) des erreurs par adjonction dans les constructions du type : **andiamo shopping* (cf. l'allemand *wir gehen einkaufen/shoppen*) → TH : *andiamo a fare shopping* ['allons faire du shopping'] ; **si può andare a trekking* → TH : *si può andare a fare trekking* ['on peut aller faire de la randonnée'] ;
- (2) des erreurs d'omission de l'article avant les adjectifs possessifs dans les FS communicatives, en particulier dans les formules de congé : **tua Lisa* (cf. l'allemand *deine Lisa*) → TH : *la tua Lisa*['ton amie Lisa] ; **tuo amico Georg* (cf. l'allemand *dein Freund Georg*) → TH : *il tuo amico Georg*['ton ami Georg'] ;
- (3) ou des erreurs d'expansion dans les formules de congé : **Con cari saluti/*Con tanti saluti* (cf. l'allemand *Mit freundlichen Grüßen, Mit vielen Grüßen*) → TH : *Cari/Tanti saluti* ['Mes amitiés'].

6. Conclusion

Le modèle de recherche phraséologique que nous avons décrit dans cet article se caractérise par des éléments novateurs qui méritent d'être mentionnés dans nos conclusions. Tout d'abord, il y a une forte composante interdisciplinaire qui combine les ressources de la linguistique de corpus, de la linguistique computationnelle et de la théorie phraséologique. Deuxièmement, c'est la phraséologie multilingue qui est au centre de l'étude, avec un accent particulier sur les productions linguistiques d'un groupe d'apprenants bien identifiés (géographiquement et sociolinguistiquement), grâce à la présence de données extralinguistiques. Le projet LeKo, que nous qualifions d'*intégral* parce qu'il rassemble des approches complémentaires pour analyser le phénomène phraséologique dans son ensemble, est donc un modèle méthodologique et procédural de recherche et de comparaison phraséologique entre différentes langues⁵².

⁵² Pour d'autres aspects du projet voir aussi: Konecny *et al.* (2016), Zanasi *et al.* (2016), Autelli *et al.* (2017), Konecny *et al.* (2019), Schmiderer *et al.* (2020), Schmiderer *et al.* (à paraître).

Enfin, les résultats du projet (tels que l'utilisation de phrasèmes communicatifs déjà détectés aux niveaux initiaux du CECRL, l'analyse des erreurs et le poids du transfert inter-linguistique) forment la base pour une phraséodidactique⁵³ ciblée.

Bibliographie

ABEL, Andrea/KONECNY, Christine (à paraître), "Formelhafte Sequenzen im LeKo-Lernerkorpus : vom theoretischen Ansatz zur Ergebnisauswertung", in Konecny Christine/Autelli Erica/Zanasi Lorenzo/Abel Andrea (eds.), *Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext, Teil 1: aus Perspektive der Semantik, Wortbildung, Translationswissenschaft, Text- und Fehlerlinguistik*, Tübingen, Stauffenburg.

ABEL, Andrea/VETTORI, Chiara/WISNIEWSKI, Katrin (2012) (eds.), *Gli studenti altoatesini e la seconda lingua : indagine linguistica e psicosociale./Die Südtiroler Schülerinnen und die Zweitsprache : eine linguistische und sozialpsychologische Untersuchung*, Vol. 1 – Bd. 1, Bolzano/Bozen, Eurac.

ABEL, Andrea/WISNIEWSKI, Katrin/NICOLAS, Lionel/BOYD, Adriane/HANA, Jirka/MEURERS, Detmar (2014), "A Trilingual Learner Corpus Illustrating European Reference Levels", *Ricognizioni. Rivista di lingue, letterature e culture moderne*, 1(2), 111-126.

AUTELLI, Erica/KONECNY, Christine/ZANASI, Lorenzo/ABEL, Andrea (2017), "La fraseodidattica dell'italiano fra continuità e innovazione : metodi tradizionali e scenari futuri", in De Giovanni Cosimo (ed.), *Fraseologia e paremiologia : passato, presente e futuro*, Milano, FrancoAngeli, 451-462.

BALBONI, Paolo E. (2012), *Le sfide di Babele*, Torino, UTET.

BURGER, Harald (2003), *Phraseologie. Eine Einführung am Beispiel des Deutschen*, Berlin, Schmidt.

BURGER, Harald (2007), "Semantic aspects of phrasemes", in Burger Harald/Dobrovolskij Dmitrij/Kühn Peter/Norricks, Neal R. (eds.), *Phraseologie/Phraseology. Ein internationales Handbuch der zeitgenössischen Forschung/An International Handbook of Contemporary Research*, Berlin/New York, De Gruyter, 90-109.

CICALESE, Anna (1995), "I composti polirematici con struttura N a N", in D'Agostino Emilio (ed.), *Tra sintassi e semantica : descrizione e metodi di elaborazione automatica della lingua d'uso*, Napoli, Edizioni scientifiche italiane, 329-349.

53 Pour des études de phraséodidactique voir aussi González-Rey (2013, 2015) et Sułkowska (2013).

- DE MAURO, Tullio/VOGHERA, Miriam (1996), "Scala mobile. Un punto di vista sui lessemi complessi", in Benincà Paola/Cinque Guglielmo/De Mauro Tullio/Vincent Nigel (eds.), *Italiano e dialetti nel tempo. Saggi di grammatica per Giulio C. Lepschy*, Roma, Bulzoni, 99-131.
- DÍAZ-NEGRILLO, Ana/FERNÁNDEZ-DOMÍNGUEZ, Jesús (2005), "Error Tagging Systems on Learner Corpora", *RESLA*, 19, 83-102.
- DOBROVOL'SKIJ, Dmitrij (2002), "Phraseologismen in kontrastiver Sicht", in Cruse D. Alan/Hundsnurscher Franz/Job Michael/Lutzeier Peter Rolf (eds.), *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*, Berlin/New York, De Gruyter, 442-451.
- FLEISCHER, Wolfgang (1997), *Phraseologie der deutschen Gegenwartssprache*, Tübingen, Niemeyer.
- GARCÍA-PAGE SÁNCHEZ, Mario (1996), "Más sobre la comparativa fraseológica en español", *Linguística Española Actual*, 18(1), 49-77.
- GARCÍA-PAGE SÁNCHEZ, Mario (1999), "Aspectos lingüísticos de la 'comparación estereotipada' en español e italiano", in Trovato Salvatore C. (ed.), *Proverbi, locuzioni, modi di dire nel dominio linguistico italiano. Atti del I. Convegno di Studi dell'Atlante Paremiologico Italiano (API), Modica, 26-28 ottobre 1995*, Roma, Il Calamo, 85-95.
- GONZÁLEZ-REY, María Isabel, (ed.) (2013), *Phraseodidactic Studies on German as a Foreign Language./Phraseodidaktische Studien zu Deutsch als Fremdsprache*, Hamburg, Dr. Kovač.
- GONZÁLEZ-REY, María Isabel, ed. (2014), *Outils et méthodes d'apprentissage en phraséodidactique*, Fernelmont (Belgique), EME Editions.
- HAUSMANN, Franz Josef (2004), "Was sind eigentlich Kollokationen?", in Steyer Kathrin (ed.), *Wortverbindungen – mehr oder weniger fest*, Berlin/New York, De Gruyter, 309-334.
- KONECNY, Christine/ABEL, Andrea/AUTELLI, Erica/ZANASI, Lorenzo (2016), "Identification and Classification of Phrasemes in an L2 Learner Corpus of Italian", in Corpas Pastor Gloria (ed.), *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Full papers)*, Geneva (Switzerland), Tradulex, 533-542.
- KONECNY, Christine/AUTELLI, Erica/ZANASI, Lorenzo/ABEL, Andrea (2018), "«*Queste vacanze divantano il martello! Transferphänomene beim Gebrauch formelhafter Sequenzen im Italienischen seitens deutschsprachiger L2-Lerner/innen und Möglichkeiten ihrer Klassifizierung",

- in Zybatow Lew/Petrova Alena (eds.), *Sprache verstehen, verwenden, übersetzen. Akten des 50. Linguistischen Kolloquiums, Innsbruck, 03.-05.09.2015*, Frankfurt a.M. etc., Peter Lang, 115-126.
- LÜDELING, Anke/WALTER, Maik/KROYMANN, Emile/ADOLPHS, Peter (2015), "Multi-level Error Annotation in Learner Corpora", in *Proceedings of Corpus Linguistics 2005*, Birmingham, University of Birmingham.
- MÜLLER, Christophe/STRUBE, Michael (2006), "Multi-Level Annotation of Linguistic Data with MMAX2", in Braun Sabine/Kohn Kurt/Mukherjee Joybrato (eds.), *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*, Frankfurt a. M. etc., Peter Lang, 197-214.
- PALM, Christine (1995), *Phraseologie. Eine Einführung*, Tübingen, Narr.
- SCHMID, Helmut (1994), "Probabilistic Part-of-Speech Tagging Using Decision Trees", in *Proceedings of the International Conference on New Methods in Language Processing*, Manchester (UK).
- SCHMIDERER, Katrin/ZANASI, Lorenzo/KONECNY, Christine/AUTELLI, Erica (2020), "Sviluppare la competenza lessicale e fraseologica tramite i *task*. Un contributo allo sviluppo di materiale per l'italiano L2", *Italiano Lingua Due 2020/2* (volume monografico a cura di Serena Enrico, Bernhard Gerald, Gallerani Irene, Visser Judith: *L'italiano in contesti plurilingui : contatto, acquisizione, insegnamento*), 238-256.
- SCHMIDERER, Katrin/ZANASI, Lorenzo/KONECNY, Christine/AUTELLI, Erica (2021), *Facciamo bella figura! 8 task fraseodidattici per studenti di italiano L2/LS*. Con una prefazione e con la consulenza scientifica di Barbara Hinger. Innsbruck: innsbruck university press. URL: <https://www.uibk.ac.at/iup/buecher/9783991060451.html>
- SCHMIDERER, Katrin/ZANASI, Lorenzo/KONECNY, Christine/ AUTELLI, Erica (à paraître), "Dal corpus LeKo ai materiali d'insegnamento : per una fraseodidattica *corpus-based*", in Konecny Christine/Autelli Erica/Zanasi Lorenzo/Abel Andrea (eds.), *Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext, Teil 2: aus Perspektive der Korpus- und Fachsprachenlinguistik, Interkulturalitätsforschung und Phraseodidaktik*, Tübingen, Stauffenburg.
- SIMONE, Raffaele (1997), "Esistono verbi sintagmatici in italiano?", in De Mauro Tullio/Lo Cascio Vincenzo (eds.), *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche. Atti del convegno internazionale della Società di Linguistica Italiana. Madrid, 21-25 febbraio 1995*, Roma, Bulzoni, 156-170.

- SKYTTE, Gunver (1988), "Phraseologie/Fraseologia", in Holtus Günter/Metzeltin Michael/Schmitt Christian (eds.), *Lexikon der Romanistischen Linguistik : Vol. IV. Italienisch, Korsisch, Sardisch/ Italiano, Corso, Sardo*, Tübingen, Niemeyer, 75-83.
- STEIN, Stephan (2007), "Mündlichkeit und Schriftlichkeit aus phraseologischer Perspektive", in Burger Harald/Dobrovolskij Dmitrij/Kühn Peter/Norrick Neal R. (eds.), *Phraseologie/Phraseology. Ein internationales Handbuch der zeitgenössischen Forschung/An International Handbook of Contemporary Research*, Berlin/New York, De Gruyter, 220-236.
- SUŁKOWSKA, Monika (2013), *De la phraséologie à la phraséodidactique. Études théoriques et pratiques*, Katowice, Wydawnictwo Uniwersytetu Śląskiego.
- VOGHERA, Miriam (1994), "Lessemi complessi : percorsi di lessicalizzazione a confronto", *Lingua e stile*, 29, 185-214.
- VOGHERA, Miriam (2004), "Composizione : Polirematiche", in Grossmann Maria/Rainer Franz (eds.), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 56-69.
- WRAY, Alison (2002), *Formulaic Language and the Lexicon*, Cambridge, Cambridge University Press.
- WRAY, Alison/PERKINS, Michael R. (2000), "The Functions of Formulaic Languages : an Integrated Model", *Language et Communication*, 20(1), 1-28.
- ZANASI, Lorenzo/KONECNY, Christine/AUTELLI, Erica/ABEL, Andrea (2016), "L'insegnamento dei fraseologismi nell'italiano come lingua seconda : proposta di un modello didattico misto", in De Marco Anna (ed.), *Lingua al plurale : la formazione degli insegnanti*, Perugia, Guerra, 171-180.
- ZELDES, Amir/RITZ, Julia/LÜDELING, Anke/CHIARCOS, Christian (2009), "ANNIS : A Search Tool for Multi-layer Annotated Corpora", in Mahlberg Michaela/González-Díaz Victorina/Smith Catherine (eds.), *Proceedings of Corpus Linguistics, CL2009*, Liverpool, University of Liverpool (UK).

Sitographie

- KONECNY, Christine/AUTELLI, Erica (2018), *Collocazioni italiane. Combinazioni di parole italiane e tedesche a confronto. Un progetto di ricerca*, <http://www.kollokation.at/it/>, page consultée le 3 avril 2018.

Profils bio-bibliographiques

Lorenzo ZANASI est chercheur à l'Institut de linguistique appliquée d' Eurac Research (Bolzano/Bozen), où il s'occupe de projets dans le domaine du développement des compétences linguistiques dans le monde scolaire et professionnel et du multilinguisme. Linguiste formée à l'Université de Rome La Sapienza et à l'Université pour étrangers de Sienne, il est également enseignant d'italien langue étrangère. A ce titre, il a exercé une intense activité professionnelle en France, au Maroc et au Sri Lanka et collabore actuellement à la Société Dante Alighieri.

Christine KONECNY est maître de conférences en linguistique italienne dans le département de Romanistique de l'Université d'Innsbruck. Depuis 2015, elle est titulaire de l'habilitation universitaire pour l'enseignement de la linguistique italienne. Ses recherches portent sur la sémantique, la lexicologie, la lexicographie et la phraséologie, où elle est considérée comme une experte dans le domaine de la recherche sur les collocations.

Erica AUTELLI est chercheur et chef de projet (www.romanistik-gephras.uibk.ac.at) aux départements de Romanistique et de Traducteurs et Interprètes de l'Université d'Innsbruck. En outre, elle a également enseigné l'italien et l'anglais pendant plusieurs années dans des collèges et des lycées et dans divers établissements d'enseignement pour adultes. Ses principaux domaines de recherche sont la phraséologie, la lexicographie, la sémantique, la dialectologie, l'enseignement des langues étrangères et la traductologie.

Andrea ABEL dirige l'Institut de linguistique appliquée d' Eurac Research (Bolzano/Bozen). Titulaire d'une habilitation en langue et littérature allemandes (2014), depuis 2005 elle enseigne dans les universités de Bolzano, Trente, Innsbruck, Ratisbonne et Mannheim. Ses domaines de recherche comprennent la modélisation de la compétence linguistique et des variations linguistiques (allemand L1 et L2), la recherche sur les corpus d'apprenants et de variétés, la lexicographie électronique et les aspects sociolinguistiques du contact linguistique.